Potential Computer Vision Technologies for Monitoring Shared Spaces

(Bristol Case Study)

Sam Gunner and Dima Damen

Technical Report
v. 1.0 May 2016

Contents

Abstract

In this report we explore the current hardware technologies capable of providing computer vision based monitoring and potential incident detection in urban environments. The required system should be able to distinguish between different moving entities (pedestrian, cyclist, vehicle) as well as detect and log interactions and potential conflicts within a shared space. We investigate a variety of options for such a system, including COTS (Commercial Off-The-Shelf), leading industrial research as well as flexible platforms currently under development for other urban sensor projects. After the introduction in Section 1 we will detail these options in Section 2. In Section 3 we case study the different technologies by considering some potential deployment locations within the city of Bristol, taking into account the detection requirements for those sites, before providing conclusions and recommendations in Section 4.

1 Introduction

Shared spaces are urban public locations that aim, by design, to minimise the segregation between moving entities (e.g. pedestrians, cyclist and vehicles). Shared spaces have tremendous advantages in saving space, increasing users' understanding of the requirements of other user types as well as strengthening communities. However, some shared spaces seem to work better than others. Several reasons have been reported for the variability in people's acceptability of a shared space. It is reported that interventions and modifications can potentially increase the 'likeability' of a shared space by its users. Analysis and interventions require data gathering. Computer vision offers an opportunity to automatically perform analysis before, during and after any interventions are introduced. This report lays the foundations for how monitoring can be achieved using available technologies.

The aim is to identify a system that will be capable of monitoring users of shared spaces. It is hoped that vision-based sensors could be easily and quickly deployed in a location where space is shared. Simple, low cost deployment is envisaged, aiming to quickly achieve information gathering about the shared space such as the nature and frequency of any potential conflicts. This report assumes that these sensors are primarily temporary. After data collection is achieved the sensor will be removed, and possibly relocated to a different site, while the information it has gathered is used to improve the design of the urban environment and hopefully alleviate any concerns of the space users. If, at a future date, interventions are made, the sensor can be reinstated, and monitoring can resume.

2 Potential Technologies

There are a large number of manufacturers and developers currently in the market delivering computer vision based analytical systems. These are however concentrated around a number of specific applications, namely traffic monitoring for road authorities and pedestrian detection for retail and security purposes. The technology used varies between either intelligent cameras, (where the video is processed on the camera hardware so only logs and event notifications are sent back over the network), to streaming the full video over a dedicated link to a central server that will perform the processing. The latter reduces the cost of the sensor itself but requires a permanent high bandwidth connection between the camera and a central location.

The majority of these COTS (Commercial Off-The-Shelf) systems are designed for a specific application, and do not provide the required flexibility in neither their video detection options nor the possible telecommunications with which they will operate. The most promising exception to this is VCA Technology, detailed below. A complete list of all COTS suppliers researched for this report can be found as Appendix 1.

Alongside established companies, there are a few smaller businesses currently developing computer vision for applications such as smart homes. One such company is Apical who, although not presently having a computer vision product on the market, are developing image processing hardware with apparently highly sophisticated functionality. More information is given below.

As an alternative to the COTS products, an option that provides high flexibility is to take advantage of recent advances in low cost single board computers (SBCs). The most famous SBC is the Raspberry Pi, however there are a large number of alternatives ranging in price and performance. Modern SBCs are only slightly larger than a credit card, allowing them to be installed at roadside, but are able to function much like a desktop PC or laptop. A number of different operating systems have been developed for these machines, but the most popular are based on the widely used Linux distribution Debian. Debian is incredibly flexible, allowing the SBC to work like a normal computer with a monitor, keyboard and mouse, or giving the option for it to be accessed remotely using a suite of secure protocols. The popularity of Linux means that a huge number of software libraries and code solutions are readily available. By connecting a visual sensor to an SBC it can be turned into a flexible intelligent camera on which new computer vision applications can be developed and run, bespoke to the task of monitoring shared spaces.

Both SBC hardware and the Linux operating systems are readily available, and so it would be possible to buy the component parts and develop a roadside computer vision system from scratch. However extra considerations such as the development of ruggedised casing, as well as the debugging process that goes with any development, means that it is preferable to use a platform that has already been developed. One such platform is the Chicago based Array of Things (AoT) Node, currently under development by the US organisation Argonne National Laboratory and based on the Hardkernel ODROID-UX4 and C1+ single board computers. The development work already achieved on the AoT node might allow for a roadside sensor to be developed more easily than if the work was started from scratch, and would link the shared space project to the wider Smart City Community. A complete list of the SBCs researched for this project is available in Appendix B.

2.1 Communications

An extra consideration when designing a system such as this is how roadside equipment will communicate its data. As touched on above, the communication requirements are heavily dependent on whether image processing is performed on the camera hardware at roadside, or by a server located away from the camera, possibly in a lab or office. If a passive camera is installed, all processing needs to be done centrally, and a high bandwidth permanent connection will be required to allow the continual streaming of video. If, on the other hand, the camera/node can perform computer vision processing itself then only occasional transmission of logs or notifications, which might include selected still images or short video clips, is required. This significantly reduces the network burden. If the intelligent camera has sufficient data storage then it is possible that events can simply be stored on board the device and collected periodically by attending site, thus preventing the requirement for connectivity at all. This sort of operation is often referred to as a standalone mode.

Although there are advantages in a device being able to operate with limited or no connectivity it is worth highlighting that, if available, a high bandwidth connection can still provide significant benefit. This is especially true during validation and testing when to confirm a computer vision system is not missing events the video is typically watched and compared with the computer vision results. The configuration of the computer vision algorithm is then modified to try and limit the number of missed events. Having a live video stream from a camera simplifies this process as the effects of any corrective adjustments made to the configuration of the computer vision can be witnessed in real time.

The provision of a live video stream to a central server also provides an opportunity for the development and testing of more sophisticated computer vision applications. Independent of the intelligence of a roadside camera, it is advantageous for a representative video stream of a shared space to be made available to a powerful computer server. This server will then be able to provide a platform on which many new computer vision algorithms can be designed, one of which could be specifically for detecting conflicts in shared spaces. This could be done in parallel with any processing on the camera hardware, and would provide an opportunity to improve understanding of what is required for a bespoke shared space monitoring system.

Commonly, a CCTV installation will have a dedicated communications link that allows the continual streaming of video back to a control room. With digital cameras, such as those discussed here, this connection would most usually be a dedicated physical IP link which would connect to the camera with copper Ethernet. However the provision of a dedicated IP connection capable of streaming uninterrupted video is extremely expensive and can have lead times of many months making it unfeasible for the temporary, flexible installations envisaged here. Often for roadside deployment the only connection available at a reasonable cost and with a reasonable lead time is that of a 3G or 4G modem. This would have sufficient capacity for the transmission of statistics, transfer of images, and if using 4G might even be able to transmit a high quality video stream. However it is important to consider that, as with mobile phones, the cost of the communications is dependent on the amount of data transferred, and this prevents even the regular streaming of video from being cost effective.

2.2 The Bristol is Open Network

A deployment of shared space monitoring cameras within Bristol city centre might be able to take advantage of the unique opportunity provided by the Bristol is Open (BIO) Network[1]. A collaboration between the University of Bristol and Bristol City Council, BIO uses a range of different technologies to deliver connectivity across some regions within the city centre. In some locations the latest ac-type WiFi should be able to deliver sufficient bandwidth to a roadside device to allow the continual streaming of live video, even when multiple cameras are installed in one location. In locations without WiFi coverage it should be possible to use an RF mesh network for the regular sending of information, although this will probably not have capacity for continual streaming of video.

[1] http://www.bristolisopen.com/

The BIO network is however a development network, therefore its permanent stability cannot be guaranteed and neither presently can the specifics of coverage. At the time of writing this report, the WiFi signal strength within the city centre has yet to be properly measured. There would then be a need to identify in which locations the WiFi signal provides sufficient connection speed. It is however felt that in areas where it is available the BIO network provides considerable opportunity for a shared space monitoring project, especially in the early stages of deployment when a high bandwidth connection could in theory be provisioned very quickly, and allow the sort validation mentioned above. It should also be highlighted that at the time of writing this report costs for the use of the BIO network have yet to be finalised.

Outside the area covered by the BIO network expense and lead-time will likely make the continual streaming of live video impractical. Therefore for a system to be used at the majority of shared spaces in Bristol, apart from the limited number covered by BIO, it is necessary for the computer vision algorithms to be performed onboard the sensing node, preferably with the option to send a video stream when connectivity is available. The standard option would be a standalone mode.

Based on the explanation, only systems that meet these criteria, i.e. can run in a standalone mode and can send information when connectivity is available, have been considered for this report.

2.3 VCA Technology

VCA Technology[2] is a UK based company specialising in computer vision for Closed Circuit Television (CCTV) cameras. They offer a large number of different cameras for a range of applications, providing a mature product that is already being used by customers worldwide. Some of the deployments reported on their webpage are for trespass detection on Chinese railway lines and pedestrian counting at large scale events in Poland. Their on-camera computer vision software is focused on traffic and pedestrian monitoring and counting, but the programming interface is flexible enough to allow a set of criteria to be met before an event will be triggered. Their platform appears to also deliver some flexibility on the communications that can be used. Discussion suggests they have experience of deploying cameras with either a dedicated connection, providing real time video streams, or with a 3G modem where videos or still images are stored on the camera and then periodically collected by someone attending site, although the specifics of this are not known.

The computer vision software runs on embedded hardware on the camera. Not much is known about the processing hardware used, although the maturity of the product makes it likely that the specification is not cutting edge. The range of optical available from VCA technology is extensive. A more detailed description of the available cameras is given below, in general though the use of hardware designed specifically for CCTV means a wide viewing angle and outdoor rated casing. Designed for security applications, many of these cameras are also equipped with infrared illuminators that allow them to work even in poor lighting conditions.

The advantage that VCA image processing appears to have over other COTS options is the flexibility with which even a non-expert can program the computer vision element after the system is installed. There are a wide range of

[2] http://www.vcatechnology.com/

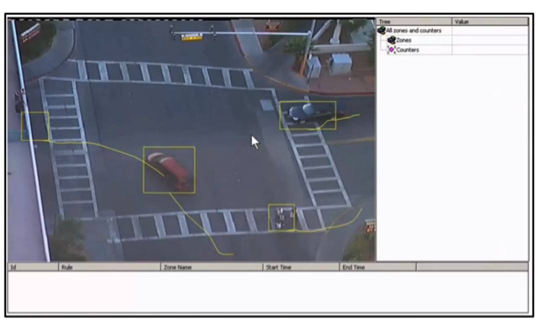


Figure 1: Screenshot of the VCA Programming Interface before configuration options have been added. Image taken from a video available on the VCA website

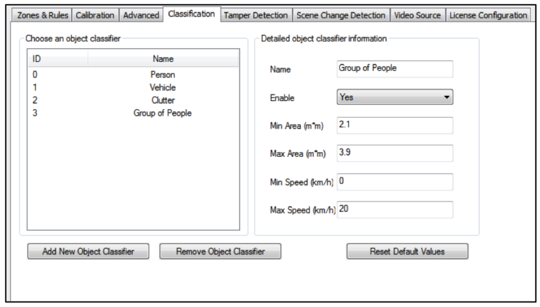


Figure 2: Screenshot from configuration interface of the VCA software, showing the default set of object classifications which can be added to, modified or deleted.

configuration options when specifying what will trigger a log entry. For the purpose of monitoring a shared space it is hoped that these options could be combined to filter out spurious information and so limit triggering to only a specific type of event. Further details on the available filters are provided next.

The camera allows different types of objects to be classified separately. Distinction between different types of object is based on the size of the object and/or the speed it is moving. It is envisaged therefore that cyclists and pedestrians could be distinguished based on their speed and cars and cyclists could be distinguished based on their size. The image recognition software includes configuration of the height and angle of installation, and using this information the camera is able to make some assumptions on the actual size of an object within the image. It should be noted that these assumptions could result in false classification (e.g. cyclist moving at slow speed).

The video image can be divided up into different zones, meaning that different rules and counts can be applied to different parts of the scene. These zones can be drawn directly on top of the live video making configuration of a zone an intuitive process. These zone will have to be reconfigured for each different deployment of the camera, and it will not be possible to do the configuration before the camera is installed and in position.

Under a default set up the camera will count all the objects of each type present within each zone. For more complicated distinction between different events, the video processing software also includes a number of filters that can be used to prevent the camera from triggering unnecessarily. Filters include: Speed, trigger only when an object is travelling between an upper and lower speed thresholds. Direction: Trigger only if an object is travelling in a specific direction, the tolerance of which is configurable. Entry: Trigger when an object enters the zone. Exit: Trigger when an object exits the zone. Appear: Trigger only if an object appears within a zone (i.e. either from a horizon or tunnel). Disappear: Trigger only if an object disappears within a zone. The camera can also trigger only

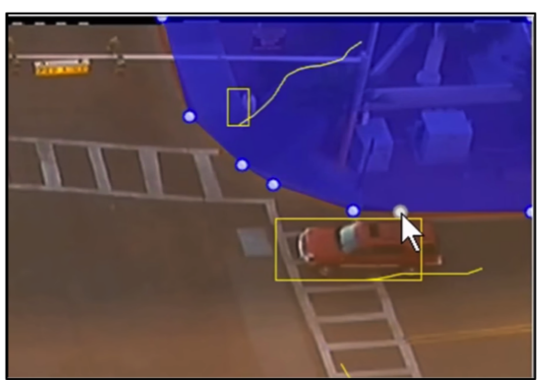


Figure 3: Zones are placed onto the live video, they can be complicated shapes and many can exist on a single camera. Image taken from a video available on the VCA website.

Figure 4: Example cameras available from VCA Technology. A motorised PTZ 'dome' camera (left) and a fixed 'bullet' camera (right). Images taken from the VCA website.

on specific event types such as dwell or tailgating. It is not known if these will be advantageous when attempting to detect conflict in a shared space.

Camera and Enclosure: The cameras used by VCA are designed specifically for CCTV applications. The optics used provide a wide viewing angle (around 100°), and the enclosures are rated for outdoor installation (IP66). Any installation or maintenance will be covered by the warranty. The only issue remaining is the potential connectivity to BIO.

A wide range of different cameras are available. Camera types include both motorised Pan-Tilt-Zoom (PTZ) cameras and static cameras (Fig 4). All of the cameras are Full High-Definition (HD), providing ample resolution for image processing and analysis, although increasing the burden on bandwidth and storage capacity. It is probable that the camera is able to transmit or store video at a lower resolution to avoid this, however this cannot be confirmed at this time. As security cameras, most of the units already have active infrared illuminators integrated. These generate infrared light that, although invisible to humans, can be seen by the camera sensor and so allows continued operation in darkness.

Communications: A number of datasheets for different VCA cameras have been examined (obtainable through the VCA website) and these give a networking interface of 100MBit Ethernet, equivalent to the Ethernet cable commonly connected into the back of a personal computer. To allow this to interface with BIO a device must be used to bridge between the Ethernet and the ac-type WiFi. WiFi to Ethernet bridges do exist on the market, however experimentation would have to be performed to ensure that it could operate effectively with BIO. Assuming this is the case it should be possible to stream video continually back over the network, although as mentioned above the coverage of the WiFi has yet to be properly tested and so specific details cannot be given.

Where WiFi is not present, a camera could connected to an RF mesh access point, and although the connection would not be fast enough to deliver a continuous video stream it could be used for the periodic sending of statistic and count values, as well as management of the camera.

Outside the BIO network it is likely that the camera will have to operate in a standalone mode. It is not entirely clear at this point how a VCA camera would function in this mode, however discussion with VCA suggests that previous installations have stored captured images locally on the camera, and these have been collected periodically by attending site. If real time count data is desired it is apparently possible to fit the camera with a 3G modem. This would not provide sufficient bandwidth for the continual transmission of video, but would be capable of delivering count information, and would allow the camera to send warnings when the internal storage starts to approach capacity.

Price: Prices given by VCA were in US dollars. The license for the management software costs \$130 while the price per camera varies with model but starts in the region of about \$250. Only one version of the management software is required, and this is used to manage and monitor the camera, as well as collate data on the events that have been counted and display video where a high speed camera connection is available.

An additional cost might be the ac-type WiFi Bridge, to allow continual video streaming over the BIO network, and these range between £50 and £200 per unit.

Pros: The maturity of the product means that very little development is required and it should be possible to get cameras installed and monitoring started quicker than other alternatives. From the relatively limited testing done so far the interface for programming the computer vision element appears to be intuitive, so that it should be possible to perform site-specific configurations in situ. A number of simple events can be detected, although it is confined to the options given above.

Cons: Although more flexible that most market ready computer vision systems, the options available for image recognition are still a limitation. It should be possible to recognise a wide range of different specific shared space transgressions, however it will not be possible to detect a more generic conflict such as a near miss between two objects because these events cannot be configured within the available set of options. Although VCA have mentioned that cameras can be used in a standalone mode, this has not been tested by ourselves and extra hardware is likely to be required.

2.4 Apical Imaging

Apical Imaging[3] is a fast growing UK based technology company currently branching into computer vision, predominantly for smart home applications. They have not yet got a complete system ready for market, but the technology they are developing appears to be the most advanced found during the research for this report, hence their inclusion. The Apical technology is based on customised processor hardware designed specifically for computer vision applications. During the preparation of this report it has not been possible to obtain many details on Apical technology. Online demonstrations show a module for tracking individuals, even when overlapped or partially obscured, in real time. The computer vision appears to be able to identify the direction of travel of individuals within an image, and identify and trigger events on specific actions such as the raising of a hand. Individuals are identified using face recognition, are tracked and the re-identified.

The image processing techniques are more sophisticated than those used on the VCA camera, and should this sort of technology be deployed for roadside CCTV it might be possible to gain real substantive information about interactions between individuals.

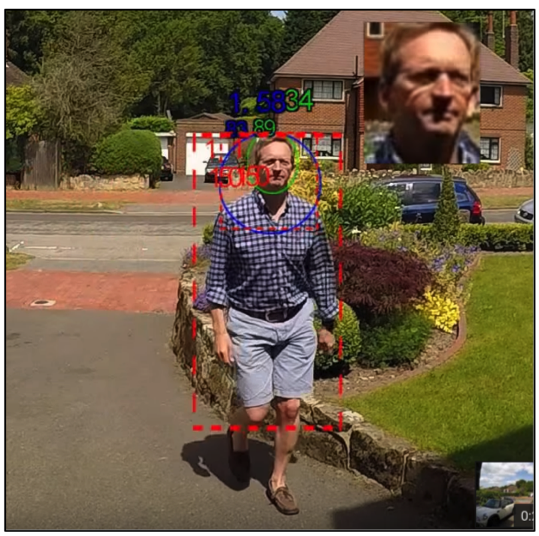


Figure 5: An intelligent doorbell application that uses computer vision to identify the presence of a person, and then facial recognition to pick out the individual from a list of contacts. Image taken from a video available on the Apical website.

[3] http://www.apical.co.uk/

2.5 Array of Things

As described above, platforms based on modern single board computers provide significant opportunity for the development of shared space monitoring computer vision applications. One of the most promising of these platforms is The Array of Things (AoT) node currently under development at the Chicago based Argonne National Laboratory, a science and research lab run jointly by the University of Chicago and the US State Department of Energy[4]. The unit is soon to be deployed across the city of Chicago, with 50 nodes being installed in 2016, and a further 450+ over the following year and half. The aim is to produce a system for monitoring in real time a large number of environmental variables including temperature, humidity, pressure, a range of different pollution levels, light levels and surface water coverage.

The AoT node does not currently perform any computer vision based detection of road users, but the hardware on which the node is developed makes it a very suitable launch pad on which computer vision based Smart City infrastructure could be developed. The use of a platform like this helps to ensure that hardware development work is not being duplicated across different IoT projects, and aids in integrating the shared space monitoring project with wider Smart Cities initiatives.

A list of the hardware being used in the AoT node is available on the project's website and Figure 6 below, showing the configuration of the node, is taken from there. The website promises that a complete specification for the hardware will eventually be available, but at time of writing this has yet to be released.

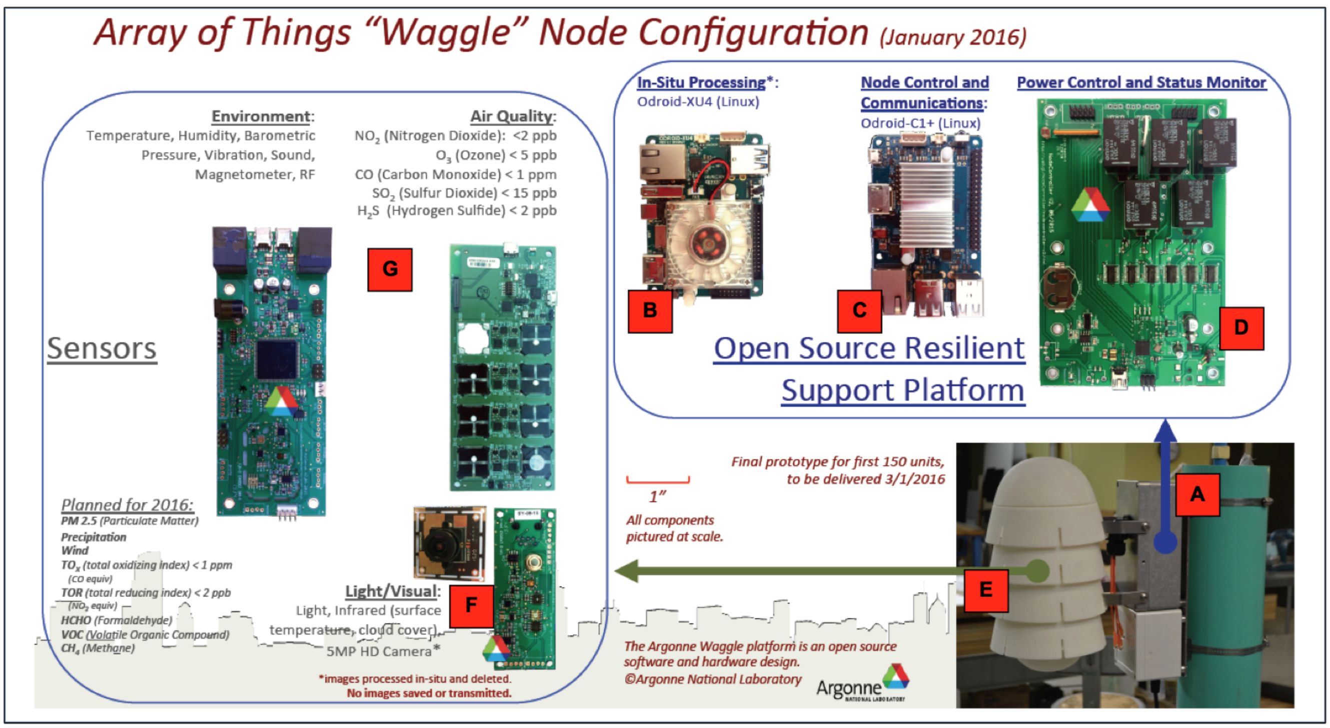


Figure 6: Hardware configuration of the Array of Things Node. Image taken from the Array of Thing website.

Figure 6 shows the modular design of the AoT node, where sensing technology is separated from the processing and communications hardware. Referencing the Figure annotations highlighted in red:

A. The "Open Source Resilient Support Platform" is where the intelligence of the unit is housed, containing the two Odroid SBCs and the Power Control and Status Monitor board.

[4] https://arrayofthings.github.io/

B. The Odroid-XU4 is the more powerful of the two SBCs on the node, this is where image processing takes place, as well as any other processing of sensor information. It is this board that will likely run any applications developed for shared space monitoring.

C. Networking is managed on the seconds SBC, the Odroid-C1+. The use of two separate processors is probably designed to increase the reliability of the platform, as well as ensure resources on the faster XU4 are not blocked by delays in the network. A common problem when running any processing device in a remote location is the threat of system crash. A crash of the operating system means connection to the device is lost making it impossible to perform remotely the reset necessary to restart operation. This results in site having to be visited to allow the device to be physically reset, and this can quickly become unmanageable if crashes happen regularly. The use of two separate SBCs within each node lowers the threat of this as if one SBC fails the other can still be contacted and can potentially perform the reset of the crashed system. Also referred to as the 'Node Control' board it is likely that the Odroid-C1+ also controls power to the other sensors installed on the node.

D. The function of the 'Power Control and Status Monitoring' board is not known.

E. The node sensors are housed separately from the SBCs in a different enclosure. This is probably to ensure that the access to the external environment required by the sensors does not negatively impact the operation of the processing hardware, and may also aid modularity.

F. The sensor of interest to this report is the visual sensor, described as a 5 Megapixel HD camera. It has not been possible to verify the exact details of the camera but more information is given on the suspected camera model, and provided next.

G. The remaining sensors provide environmental information. These are beyond the scope of this report.

As can be seen from the table in Appendix B (relevant information repeated in Fig 7) the Odroid-XU4 is one of the more powerful SBCs currently available. It has sufficient processing power to perform complicated computer vision analysis, and videos are available on the Hardkernel website showing some applications that have been developed so far[5].

Name	Odroid-XU4
Processor	A15 2Ghz and Cortex^{TM}-A7 Octa core CPUs
RAM	2GB
RAM Bus	933MHz
Onboard Networking	Gigabit
Onboard Storage	Option of 64GB

Figure 7: A photo of the Odroid-XU4 taken from the Hardkernel website, alongside the technical specifications of the board.

Camera: As mentioned, amongst the sensor hardware is a 5 megapixel HD camera. After discussion with Argonne it is believed that the node includes this specifically for the purpose of measuring surface water coverage, although there is also mention on the website of using the colour of the sky to infer pollution levels, which may require a second camera. Not much is known about how these will work but the website rules out the transmission of video stating that all image processing will be done on board the node. *Considering the applications for which these cameras have been previously used, they will most likely be positioned to look directly up, and directly down, although again this is purely speculation at this stage. If this is the case some hardware redesign is needed to enable the camera to be directed towards a point of interest such as a junction or shared space.*

There are a number of cameras compatible with the Odroid-XU4, but it appears from the information give on the website that the sensor being used on the AoT node is the 5Megapixel oCam, supplied through the Hardkernel

[5] `http://www.hardkernel.com/main/products/prdt_info.php?g_code=G145231889365`

Figure 8: The oCam, supplied through the Hardkernel website and a potential camera for the AoT node. Image taken from the Hardkernel Website.

website. This camera connects to the SBC via USB3 and is described as 'plug and play' meaning it should require no extra software or drivers to be installed if the Odroid-XU4 is running the supplied operating system.

The angle of view of the camera is 65°, comparable to most smartphones. This is better than is achieved with cheaper SBC cameras such as the one commonly used with the Raspberry Pi, but significantly worse than that of VCA cameras. It is not known if the lens on the oCam can be modified to increase the field of view as it can be on the Raspberry Pi. While more coverage can be achieved by installing the node higher above the ground, this will affect the resolution of the camera, affecting the level of detail in the received information. As an example of the impact of this angle of a view, a camera with a field of view of 65° positioned 10m away will have a horizontal viability of 12.7m, this is compared to a camera with a field of view of 100° which will have visibility of 23.8m at the same distance.

On Board Computer Vision Application: Argonne has confirmed that their development does not include any image recognition to detect the presence of pedestrians. As mentioned above however the processing hardware used in the sensor, the Odroid-XU4, is capable of performing quite sophisticated computer vision, and the real advantage of a Linux based platform is the ease with which applications can be developed, tested and installed. This development will build on a range of software libraries already available for Linux that perform basic computer vision functionalities (e.g. Open Source Computer Vision (or OpenCV)[6] - with confirmed operation on Odroid-XU4). A new interface can offer flexibility for custom building of functionalities suitable for public space monitoring purposes.

AoT provides an opportunity to develop specialised software to perform intelligent analysis of the image, such as estimating time to collision or detecting near misses in shared spaces. The use of the AoT sensor platform would allow for the development of a truly bespoke conflict detection system designed specifically for the task.

The drawback of this development would be the time required to develop this novel software, slowing the speed with which a system could be deployed. The innovative nature of this work makes it extremely difficult to know how long this development would take. It must also to be highlighted at this stage that, although the Odroid-XU4 is one of the higher specification SBCs on the market, computer vision is a resource intensive set of algorithms. The more sophisticated the computer vision used the more processing power required, and this puts an upper limit on the amount of image processing that can be performed in real time. It is not possible to know exactly what level of real time image processing will be possible on the Odroid-XU4 prior to experimentation.

Centralised Computer Vision Application: In the documentation about the AoT node Argonne do state that in their deployment the node will not be used for the streaming of live video. However this appears to be entirely for reasons of citizen acceptance and not a technical limitation, meaning that a deployment in Bristol could perform this functionality, under certain circumstances, if it is was felt that the public would not object.

[6] `http://opencv.org/`

Figure 9: The Odroid-XU4 performing face detection using the OpenCV software library. Full video can be found on the Hardkernel website.

The advantages of performing the computer vision centrally would be the significant increase in processing resource available. Servers considerably more powerful than the Odroid hardware could be used, which might enable even more sophisticated computer vision to be performed. However, as highlighted before in this report provisioning a connection capable of delivering live video will probably not be possible in every location where sensor deployment might be sort, and so it would still be necessary to develop on board image processing applications for these locations.

One potential advantage of using the flexible AoT node for roadside computer vision is the possibility of performing intelligent compression before broadcasting transmission back to a server for processing. It might be possible, for example, to convert the video to an array of direction vectors with a low enough bandwidth requirement to be transmitted over legacy WiFi, but still contain the information required by the central server to detect conflict events. This allows the sharing of the video processing burden between a roadside devices and a central server, and would better address privacy concerns as the original video would not be transmitted.

Communicatoins: Like all Linux based systems, the AoT node can be adapted to use a variety of communication methods, or it could operate as a standalone system with limited or no connectivity. Discussions with Argonne suggest that the current deployment of the sensor uses campus WiFi for its communication requirements. Future deployments outside the reach of this WiFi network will use 3G modems, but these are not yet available. It is not known exactly the sort of WiFi that is being used in the AoT node, but it is assumed to be 802.11b, g or n as these are most common.

Argonne clarify that if a user has the expertise to adapt the node to use a different communications technology then doing so is at the discretion of the user. However they do strongly advise against this while the node is still being debugged, and they will not be able to provide support or assistance with upgrades for nonstandard devices. For the purposes of this report we will consider the full flexibility of the platform.

Within the area of the BIO network it should be possible to use the BIO WiFi access points. The WiFi already on board the AoT node could be used for the polling of information, downloading of images or even streaming relatively low quality video. However the streaming of video over legacy WiFi connections, especially in areas with large amounts of radio frequency noise such as busy urban environments, can be unreliable and prone to dropouts. It would be possible to upgrade the WiFi in the AoT node to the more modern ac-type. This newer standard is able to deliver considerably higher bandwidth and should provide better resilience to background noise. BIO is equipped with this type of WiFi, but as mentioned is only planned for constrained locations. Other, lower bandwidth, connection media offered by BIO could be used such as RF mesh. This would not be able to provide the transport of live video, but would be able to deliver notification, statistics and still images, as well as sufficient data for the management of the sensor such as performing software updates.

If communications is desired outside of the BIO network the most suitable technology would be 3G. The AoT node could be fitted with a 3G modem that would allow the periodic transfer of statistics. This would require a SIM card and, as with a mobile phone, the price of the communications is based on the amount of data transmitted. It would be possible to transmit photos via 3G, but this would add considerable operational costs to the platform. Instead it is preferable to send only minimal data, saving images or video to storage onboard the node and transmitting limited statistics such as a count of the number of witnessed incidents.

Again, the flexibility of the Linux based platform means that the applications developed could function almost entirely standalone with limited contact with a network. It is not known what onboard storage the AoT sensor is currently fitted with, however the Odroid-XU4 is compatible with eMMC modules that allow the addition of up to 64 gigabytes. This would be capable of storing about twenty hours of HD video, or many thousands of hours at a lower quality suitable for validation. The computer vision application could be configured to save video and images to this internal storage, rather than transmit it across the network, and that could be collected from the device periodically, or when a warning is received as the internal memory becomes full. Depending on the length of time a camera is deployed for, and the number of incidents expected, this amount of storage might be sufficient to prevent the requirement for periodic checking of the sensor. In standalone mode, the node could act a secure WiFi access point, allowing a site attendant to connect to it and download all the images and videos automatically without requiring a physical connection to the box.

Price: Since the AoT node is still under development an exact price per unit cannot be given. The AoT website gives a cost of anywhere between \$500 and \$2000 for each node depending on the amount of sensing technology installed. The price for the oCam visual sensor given on the Hardkernel website is \$99.

Other costs associated with using this platform include, but are not limited to: the cost of developing the computer vision application that would run on the AoT node, hardware development involved in repositioning the camera on the node to allow the monitoring of shared space, the cost of the 3G contract for devices outside the BIO network, potential costs involved in using the BIO network.

Pros: The primary advantage of using a platform like the AoT node is the flexibility it offers. This would allow the development of cutting edge computer vision applications designed specifically for detecting conflict between shared space users. This is a significant distinction from the COTS alternatives. The pros when comparing the use of the AoT node with developing a system completely from scratch are the reduction in required development, the prevention of the duplication of effort between different Smart City projects, as well as providing an opportunity to concrete relationships with the wider Smart City Community.

Cons: As described above, the drawback of using this sort of platform is the significant development that would be required to produce a device that is able to deliver the required functionality. The AoT node does not currently perform any detection of this sort, and so a computer vision application would have to be developed and tested before any sensors could be deployed. The hardware would also have to be adapted to position the camera in a way that allows the monitoring of shared space.

2.6 Other Considerations:

Power: Both the VCA camera and the AoT node will most likely require a permanent power supply. Although it would be possible to run either of these systems from a battery, to do this for anything more than a few hours would require a battery with a capacity considerably higher than that provided by a normal mobile phone or laptop battery. Batteries similar to car batteries can be used, and are relatively inexpensive, but their large volume and weight can prevent them being mounted at height and this may add complications when considering the installation of the system.

Solar panels can be used to extend the life of a battery, but again these are expensive and the requirement for direct sunlight to achieve a responsible power output can increase the complexity of an installation and limit the locations where they can be used.

For a completely dependable power supply, it is preferable for the device to be connected directly to the 230V mains. If a camera or sensor node is installed on a lamppost or on existing CCTV installation it might be possible to share the power supply already provisioned for that, although some sort of agreement would have to be made with the local authority to ensure that the cost of the electricity is fairly distributed. The provisioning of a completely separate power supply for an installation would be possible, but the lead time for installation is likely to be long.

Installation Requirements: Cameras will need to be installed in a suitable location, and for most applications this is at height. This provides the best viewing angle of the scene, avoids obscuring the camera simply by someone standing close by and improves the security of the installation. It is normal for a dedicated pole to be installed, however the cost of this pole and the civil engineering work involved with the installation will likely rule this out for the temporary deployments considered here. It is therefore highly preferable to have these cameras share existing infrastructure or be mounted on existing objects or buildings. This will however require agreement from the owner of the infrastructure.

Security: Anything left unsupervised at roadside is at risk of being stolen or vandalised. This is best avoided by installing the infrastructure at height, however even this will not avoid occurrences of vandalism completely and so some operational cost should be set assigned for repairs and replacement of infrastructure.

Legislation: Legislation for the installation of cameras in public places does exist in the UK, including but not limited to requirements on accompanying signage. It has not been possible for this report to discover exactly how this relates to computer vision systems, but it is expected that any system streaming live video will be subject to these constraints. Before installing cameras at roadside it is recommended that this legislation is properly examined and understood to ensure that the project is not in breach of the law. Information can be found on the UK Government website[7] relates primarily to domestic CCTV but gives links and contacts to a wider CCTV code of conduct.

[7] `https://www.gov.uk/government/publications/domestic-cctv-using-cctv-systems-on-your-property/domestic-cctv-using-cctv-systems-on-your-property`

2.7 Summary Table

	On Board Computer Vision Capability															
	Pedestrian Detection	Cyclist Detection	Vehicle Detection	Speed Detection	Object Tracking	Object Counting	Object Interaction Detection	Near Miss Detection	Video Streaming Capability (when high bandwidth link is provided)	BIO Communication Technologies Available	Standalone Storage Functionality	Communications Technologies Available Outside BIO	Camera Specification	Enclosure Specification	Price	Product Maturity
VCA Technology Image Recognition Enabled CCTV	Yes	Yes (probably)	Yes	Yes	Yes	Yes	No	No	Yes	WiFi (with bridge developed separately), REF Mesh Network (sending statistic)	Count information and images stored on the device.	3G (for sending statistic and warnings), standalone	Many Cameras available, 100° angle of view, IR illuminator, Full HD	Full outdoor rating (IP66), either fixed or PTZ	License - 130USD, camera - 250USD per unit	Very mature and used worldwide
Apical Image Recognition Hardware	Yes	Yes (probably)	Yes (probably)	Yes (probably)	Yes	Yes	Yes (probably)	Yes (probably)	Unknown	Unknown	Unknown	Unknown	Unknown	None	None	Still largely conceptual
Argonne Array of Things Node	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes (probably)	Yes	WiFi (dongle integrated into device), REF Mesh Network (for statistics and device management)	Count information, images and videos could be stored on the device. Up to 64GB storage.	3G (for stats and warnings), standalone (the possible WiFi access point to aid data collection)	65° angle of view, no IR illuminator, 5MP	Unknown, camera installation still to be developed	Node - between 500USD and 2000USD per unit	Pilot node hardware is currently being rolled out, computer vision software still to be developed.

Figure 10: Hardware and Software Capabilities for Shared Space Monitoring

3 Case Studies

For this report three case studies (CS) have been examined to analyse how the different technical options proposed in Section 2 might be deployed in real world scenarios. The locations for these case studies have been chosen from the initial responses to a social media campaign ShareBristol[8] run by the University of Bristol to identify locations within the city where it was felt the current urban design led to conflict between users of a shared space.

The sites have been chosen to highlight the sorts of location where computer vision might be used effectively, as well as places where the space and the nature of the conflict mean that it would be much harder to identify instances of a conflict using computer vision.

3.1 CS1. Temple Way Underpass

Description:	Shared Cyclist, Pedestrian Underpass
Postcode:	BS2 0BU
Lat, Long:	51.454598, -2.583288
	This is outside the proposed locations where BIO connectivity would be available.

The Temple Way Underpass is a shared space route with a central division separating cyclists and pedestrians. At each end of the tunnel traffic calming barriers have been erected on the bicycle side of the path to prevent cyclists from entering the underpass at speed. These barriers have not however been put on the pedestrian side of the path and so some cyclists feel encouraged to use the wrong side of the tunnel.

The detection of this sort of transgression is possible using computer vision techniques. As can be seen from Fig 11, the area of interest is restricted to that highlighted by the red box. The action on which an event should be triggered is also well defined, limited to when a cyclist is seen entering or exiting the tunnel on the wrong side. This reduces the complexity of the image analysis required by the computer vision.

BIO connectivity would not be possible at this location. Therefore the camera would have to operate without a permanent connection to a central server, and all image processing will have to be done on board the device.

Option 1: VCA Technology

This sort of simple and well-defined infraction is suited to the computer vision capabilities of the VCA camera. An object would have to be created that successfully identifies cyclists and distinguishes them from pedestrians. Using the knowledge that cyclists travel on average faster than pedestrian it should be possible to perform this distinction based on the speed of an object. Different thresholds could be tested until one with a sufficiently high accuracy was found.

The camera could save an image the events that caused it to trigger, and these could be stored to an internal memory that could then be collected periodically. This would allow the camera to function as a standalone device without connection to the wider network. A 3G modem could be used to send statistic and warnings to a server.

Option 2: AoT Node

For a straight forward installation like this the increased flexibility of the AoT node might not provide a significant advantage. Again the computer vision would be looking for cyclists within a specific part of the image, the difference being that the AoT node might be able to distinguish between cyclists and pedestrians using more sophisticated techniques like shape recognition, rather than based on the size and speed on an object. It is not possible to know without development and experimentation whether this would have a significant improvement on the robustness of the detection.

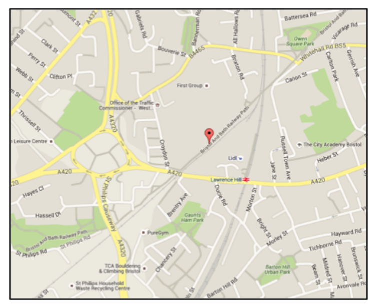
Like the VCA camera, the lack of permanent connection for the AoT node means that it would have to save all images and logs to an internal storage until they could be collected, and 3G could be used for stats and warnings.

[8] `http://tiny.cc/sharebristol`

Figure 11: Underpass entrance with highlight showing section designated for pedestrians

3.2 CS2. Bath Bristol Cycle Path at Lawrence Hill

Description:	Cycle path, heavily used by cyclists and pedestrians
Postcode:	BS5 (no specific postcode)
Lat, Long:	51.459332, -2.565646 This is outside the proposed locations where BIO connectivity would be available.



The Bath Bristol Railway Cycle Path takes cyclists and pedestrians traffic free all the way from the centre of Bath to the outskirts of Bristol City Centre. It is intensively used, especially during rush hour or weekends when large volumes of commuter and school traffic can lead to conflict along much of the path. Pedestrians are often walking two or three abreast, blocking the flow of cyclists, and cyclists are often going at speed, intimidating pedestrians. The sheer volume of traffic can lead to near misses.

Generalised areas such as this, where the conflict in question is poorly defined with a large number of different potential causes, are very difficult to interpret using computer vision even when complete camera coverage is available. As can be seen from Figure 12 (left), although a considerable length of the cycle path can be captured in a single image, this can only be done with a very shallow viewing angle, which would cause much overlapping of objects and obstruction of the rest of the scene by those objects closest to the camera. The alternative is to mount the camera looking directly down at the ground, as shown in Figure 12 (right), however by doing this we limit our field of view to a tiny area and so reduce our chances of detecting a conflict accordingly.

Option 1: VCA Technology

Combined with the complication of view, the loose description of the conflict we are trying to detect makes it impossible to identify using the VCA camera. It is not possible a convert this sort of conflict into an action that could be programmed into the VCA software. One might be able to distinguish between cyclists and pedestrians, and record the presence of each, but the large number of people using the cycle path without incident would swamp any events of interest, and none of the filters available on the VCA camera would be able to single out conflict events.

Option 2: AoT Node

A more flexible computer vision system on the AoT node might have a better chance of detecting more generic

Figure 12: The Bath Bristol Cycle Path near Lawrence Hill (left). A steep viewing angle will make distinguishing between different objects easier, but will reduce significantly the coverage area (right).

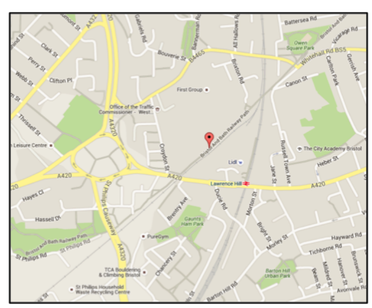
conflict by analysing the movement of objects within an image and identifying those that come close to collision. This technique however would struggle when objects are obscuring each other or are overlapping. The success of any algorithm would rely on the resolution of the image, its frame rate and the viewing angle.

However, while algorithms can be developed, it is not possible to cover the whole length of a path with a single camera independent of the viewing angle. Tens of nodes would be needed to cover sufficient length and these nodes would likely flag up a large number of events before any true conflicts were identified. It might be possible to reduce the processing needs by configuring the nodes to only operate during rush hour. The nodes can still be useful as a small experiment over a small segment of the path. This would be an experimental test and its success cannot be estimated beforehand.

It should be noted that, in general, trying to monitor an extremely large area for a poorly defined conflict type is difficult and could be error-prone.

3.3 CS3. City Centre Fountains

Description:	Shared open space with cycle path running through
Postcode:	BS1 4DA
Lat, Long:	51.452444, -2.597598
	It is intended that this location will have BIO WiFi coverage.



The Bristol City Centre Fountain area is surrounded by a large, busy, shared space. It is on most of the routes across the city, as well as being a destination in itself as it is surrounded by restaurants, bars and amenities. Pedestrians and cyclists cross the square travelling in a wide range of different directions, and this can add to the potential conflict in the area. This can be seen from the large number of complaints brought up about the area during the ShareBristol social media campaign. Complaints range from pedestrians accusing cyclists of going too fast, to more general feelings of confusion caused by insufficient road markings.

This area provides an opportunity to develop improved computer vision techniques for conflict detection while still allowing the quick rollout of a simpler system for detecting events possible to identify using existing technology. This opportunity is made available by the high speed access that should be possible through the Bristol is Open Network. This high bandwidth communication might allow the continual streaming of live video from a camera at roadside to a centralised server. This server could then be used as a platform on which new computer vision applications could be developed.

Another advantage of the fountain area is that there is already a large amount of street furniture, including BIO infrastructure and local authority CCTV that a camera or sensor might be mounted to, and powered from.

Figure 13: The shared space adjacent to the Bristol City Centre fountains.

Option 1: VCA Technology

If a VCA camera was installed within the area of the fountains it should be possible to use an ac-type WiFi bridge to connect it to the BIO network. The camera could be configured to detect cyclists based on their size and speed, and then trigger an event only on those travelling faster than a specific threshold. At the same time the high resolution video being returned over the BIO network could be used to develop more sophisticated computer vision techniques for detecting more complicated conflicts or near misses.

The relatively wide viewing angle of the VCA optics means that a wider scene could be observed than with the AoT platform, and so in that respect the chances of detecting these conflicts would be increased.

Option 2: AoT Node

Initially the AoT node would provide similar functionality to the VCA camera. The AoT sensors would still be able to stream video back to a central server for processing using an ac-type WiFi dongle. The advantage the AoT node has over the VCA camera is the potential for applications to be ported from a central server to the roadside hardware as their development progresses. The central server could be used as a platform on which applications are developed and tested, and as these application reach maturity they are adapted to allow them to operate on the embedded Linux of the Odroid-XU4. This would then mean the advanced detection mechanism afforded by the new application could be used in locations where the BIO network is not available. This highlights the real advantage over the VCA hardware, which would not allow for the development of increased functionality over time.

4 Conclusion

As demonstrated, both the VCA technology and the AoT node have potential advantages. The VCA camera provides a mature computer vision platform that can be configured out of the box to detect a range of different transgressions, whereas the AoT node provides a powerful and flexible foundation upon which novel shared space monitoring applications could be developed.

It is believed that both systems could be interfaced with the Bristol is Open (BIO) Network where available, and transmit live video back to a central server for further processing and analysis. Both could also be run standalone, or with only limited connectivity in areas where BIO is not available. In this configuration both the VCA camera and the AoT node would be able to log events, and store images on internal storage for periodic collection. It is likely that the flexibility of the AoT node would make it easier to add more storage, and develop novel way of collecting the data, but the exact functionality of the VCA camera in this respect is not known and so it is not possible make a direct comparison.

The decision on which platform to use should depend on the amount of development and technical expertise available to this project over the short and long terms, as well as the aspirations for the project as a whole.

If seen as the start of a longer term technical project the AoT node does provide a real opportunity. It would probably be possible to develop software for the AoT node able to perform similar computer vision functionality as is currently available on the VCA camera. However with further development, using some of the expertise present at the university and the opportunities presented by the BIO network, it might possible to create novel solutions, enhancing our understanding of interactions in shared environments, and progressing towards the goal of Smart Cities.

Appendices

A CCTV Suppliers

Complete list of potential suppliers of Computer Vision CCTV surveyed for this report. If contact information is required for these companies please request this from the writer of this report.

Company	Company Description	System Description	Website	Indicative Price
VCA Technology	Surrey Based Image Recognition Developer, Korean Company UDP is parent	Camera based image recognition, running on relatively old ARM9 processor. Camera can be programmed to start recording when a specific event it witnessed, these events are things such as objects of specific size and speed in a specific location. Video can either be streamed or downloaded right away if a connection is available, or can be saved to a camera for collection later. Camera can be connected to a 3G modem and stats such as number and type of events can be sent via 3G if no other connection is available. A central server is available for managing cameras. Camera image recognition software is also available on Windows.	http://www.vcatechnology.com/	software license - $130 range of camera starting from $200 per unit
Apical	London based company that produces screen and video processing technology	A number of seemingly cutting edge technologies that do image recognition, so far not much information is available from the company however information from VCA suggests that Apical are creating specific processor hardware for image recognition. Website states that camera technology is based on Soc and FPGA technologies. The website suggests that everything is done on the node. After brief discussion with the company it was found no actual product is available at present.	http://www.apical.co.uk/assertive/camera/	unknown
Argonne	National Research Lab operated for the US Department of Energy, created the Array of Things Sensor.	The Array of Things Sensor is based on two Odriod boards, the XU4 processing data, and the C1+ managing communications. The platform is used for wider data collection. A camera is onboard however the image recognition is limited to identifying ground water and city pollution (using the colour of the sky). All video is processed online and only calculated values are stored. XU4 can do more sophisticated image processing, and Argonne are working with partners to possibly develop this in the future. Communication is all currently done via WiFi, although there are plans to include 3G for future version. Platform is still under development, and debugging is still on going.	for Argonne: http://www.anl.gov/ for array of things: https://arrayofthings.github.io/	still under development so a final price for the unit is not clear and price upward of $500 is quoted. Odroid XU4 - $75 Odroid C1+ - $37
ClearView Communications	Sussex Based Security Company	Camera Based Image recognition based on size and speed of object. Technology apparently supplied by VCA technology.	http://www.clearview-communications.com/	Unknown
Axis Communications	Swedish CCTV Company	Wide range of high end IP enabled CCTV, very limited image processing and designed specifically for a secure hard wired connection.	http://www.axis.com/gb/en/	> $300 per camera
Hikvision	American IP CCTV Company	Lower budget CCTV cameras and Digital Video Record than Axis, more targeted at domestic use but again expects high bandwidth connection. No mention of image recognition on website.	http://overseas.hikvision.com/en/product_1.html	< $100 per camera
Intelligent Security System CCTV	American Security Company	Company produces a range of image recognition systems including face recognition and traffic detection. Website suggests that all processing is done at a central server as no specialist camera is required.	http://issctv.com/	
Citilog	International company providing traffic monitoring and detection image recognition systems.	Company provides a number of different products, using a few different types of image sensor. They are aimed exclusively at traffic detection of different forms, these include traffic counting, presence detection and applications such as detecting stopped vehicles and vehicles travelling on the wrong side of a carriageway. Visual sensors used are mostly Axis cameras or the XCam. It appears that the XCam is designed to interface specifically with legacy UTC systems and therefore does not output a video stream, while the Axis camera is designed to integrate with a CCTV system and will return the video as well and the data generated from the image processing.	http://www.citilog.com/en-compare	

Figure 14: List of potential suppliers of Computer Vision CCTV surveyed for this report

B Single Board Computers

Complete list of Single Board Computers surveyed for this report.

Unit	Processor	RAM	RAM Bus	Onboard Networking	No. USB	SATA	Guide Price ($)	Frequently Used Operating System	Camera Available
Raspberry Pi	700 MHz ARMv6 – Broadcom	512 SDRAM	400 MHz	100 Mbit Ethernet	4No. USB2	no	$29.99	Raspbian	yes
Raspberry Pi 2	900 MHz ARMv7 Quad – Broadcom	1GB DDR2	450 MHz	100Mbit Ethernet	4No. USB2	no	$35.00	Raspbian	yes
Raspberry Pi 3	1.2GHz Quad-core	1GB LPDDR2	Unknown	WiFi, Bluetooth, 100Mb Ethernet	4No. USB2	no	$40.00	Raspbian	yes
Raspberry Pi Zero	1GHz Single Core	512MB	Unknown	No native	1No. USB1 micro socket	no	$10	Raspbian	no
Banana Pi	1GHz ARM7 Dual	1GB DDR3	432 MHz	Gigabit Ethernet	2No. USB2	yes	$36.99	Bananian	yes
Banana Pi Pro	1GHz ARM7 Dual	1GB DDR3	432 MHz	Gigabit Ethernet, WiFi	2No. USB2	yes	$45.00	Bananian	yes
HummingBoard	i.MX6 Quad	4GB DDR3	Unknown	Gigibit Ethernet, WiFi, Bluetooth	4No. USB2	yes	$217	Android	yes
BeagleBoneBlack	1GHz	512 DDR3	Unknown	100Mbit Ethernet	1No. USB1	unknown	$60	Linux	yes
PandaBoard	1.2GHz Dual + GPU	1GB DDR2	Unknown	100Mbit Ethernet, WiFi, Bluetooth	2No. USB2	unknown	$120	Linux	yes
pcDuino3B	1GHz ARM Dual	1GB	Unknown	Gigibit Ethernet, WiFi	1No. USB2	yes	$59.99	Linux	yes
pcDuino3Nano	1GHz ARM Dual	1GB	Unknown	Gigibit Ethernet	2No. USB2	yes	$39	Linux or Android	yes
IntelGalileo	Intel Quark 400MHz	512 SRAM	Unknown	100Mbit Ethernet	1No. USB2	no	$75	Arduino Linux variant	no
Intel NUC	i5	2x DDR4 slots	Unknown	Gigabit Ethernet	4No. USB3	yes	$329	Windows 10	no
CottonCandy	Samsung Exynos 1.2GHz	1GB DRAM	Unknown	WiFi only	1No. USB2 (adapter required)	no	$199	Linux	no
Odroid-CI+	A5 1.5GHz Quad ARM + Mali-450 MP2 GPU	1GB DDR3	Unknown	Gigabit Ethernet	4No. USB2	unknown	$35	Ubuntu	yes
Minnowboard Max	Intel Atom 1.33GHz Dual	2GB DDR3	Unknown	Gigabit Ethernet	1No. USB2 1No. USB3	yes	$139	Linux, Android or Windows	yes
OLinuXIno-LIME2	Allwinner A20 dual 1GHz	1GB DDR3	Unknown	Gigabit Ethernet	2No. USB2	yes	$55	Linux	yes
Cubietruck	AllWinner A20 dual 1GHz	2GB DDR3	Unknown	Gigibit Ethernet, WiFi	2No USB2	yes	$119	Linux or Android	yes
Orange Pi	AllWinner H3 quad 1.6Hz	1GB DDR3	Unknown	Gigibit Ethernet, WiFi	2No. USB2	yes	$41.58	Linux or Android	yes
PCEngines APU	AMD G series, 1GHz	4GB DRAM	Unknown	3x Gigabit Ethernet, DB9	2No. USB	yes	$180	Linux	no
MIPS Ci20	MIPS – JZ4780 – 1.2GHz Dual	1GB DDR3	Unknown	100Mbit Ethernet, WiFi, Bluetooth	2No. USB	no	$65	Debian	yes
Odroid-XU4	Samsung Exynos5 Octa ARM Cortex - plus other options	2GB LPDDR3	933MHz	Gigabit Ethernet	2No. USB3 + 1No. USB2	no	$74	Linux or Andoird	yes

Figure 15: List of Single Board Computers surveyed for this report